Claims:

1. A method for balancing unicast or multicast flows in a multistage non-blocking fabric, wherein the fabric comprises at least one internal switching element (SE) stage, wherein the stage has *l* internal switching elements, wherein each internal switching element is associated with a unique numerical identifier, and wherein the fabric comprises an input SE stage and an output SE stage, wherein the method comprises:

(a) grouping input ports into input sets whereby each input set consists of input ports that transmit through the same input SE, and wherein the input sets are divided into input subsets, and

(b) grouping output ports into output sets whereby each output set consists of output ports that receive cells through the same output SE, and wherein the output sets are divided into output subsets,

(c) receiving cells into the fabric wherein

if a cell is a unicast cell, then the cell is associated with an input subset *i* and associated with an output subset *j* based on the input port and the output port of the cell, else

if a cell is a multicast cell, then the cell is associated with an input subset *i* and associated with multiple output subsets {*j*} based on the input port and the multiple output ports of the cell,

(d) assigning each cell to a flow, wherein

(i) if the cells are unicast cells, then the cells which are associated with the same input subset and associated with the same output subset are assigned to the same flow, else

(ii) if the cells are multicast cells, then the cells which are associated with the same input subset and associated with the output subsets of the same output sets are assigned to the same flow, and

(e) transmitting flows through the internal SE stage wherein cells of a particular flow are distributed among the internal switching elements, wherein the quantity of the cells of each particular flow transmitted through each internal SE differs by at most *h*, wherein *h* is positive, wherein the number of subsets of at least one input set or at least one output set is less than n, wherein n is the number of ports of that input SE or of that output SE, and wherein N is the total number of input ports and output ports, and wherein N_f is the maximum number of flows whose cells pass any given link, and wherein n, N, N_f, h, i, j and l are natural numbers, wherein the flow in the fabric is balanced.

2. The method according to Claim 1 wherein h is one.

3. The method according to Claim 2 wherein step (e) comprises the sub-steps of:

(a) associating a unique counter with a flow, wherein a counter is designated as c_{ij}, wherein i is the numerical identifier of an associated input subset and j is the numerical identifier of an associated output subset;

(b) initializing the counter with a number less than or equal to l;

(c) transmitting a cell from the flow through the internal switching element associated with a numerical identifier which is equal to the numerical value of the counter and

(d) changing the numerical value of the counter by decrementing or incrementing the counter modulus l, and

(e) stopping if there are no more cells from the flow, otherwise go to step (c),

wherein the sub-steps are performed for each flow.

4. The method according to Claim 2 wherein at least one of the input ports belongs to at least two input subsets, or wherein at least one of the output ports belongs to at least two output subsets, or combinations thereof.

5. The method according to Claim 3 further comprising:

grouping cell time slots into frames of length F,

wherein the counter of each flow is set at the beginning of each frame, and wherein the counter is set to c_{ij}=(i+j) mod l.

6. The method according to Claim 2 further comprising:

grouping cell time slots into frames of length *F*,

allowing in each frame input port (*i*) to transmit up to a_{ij} cells or high-priority cells to output port (*j*), and wherein:

$$\sum_k a_{ik} \le SF - N_f\,, \quad \sum_k a_{ki} \le SF - N_f$$

where *S* is the switching fabric speedup.

7. The method according to Claim 6 wherein at each stage only cells that have arrived in the same frame are transmitted to the next stage, wherein $F=D/3T_c$ or $F=D/4T_c$ if cells are reordered at the outputs, wherein *D* is the maximum tolerable delay and T_c is cell time slot duration.

8. The method according to Claim 6 wherein:

$$S = 1 + \frac{N_f}{F},$$

and: $\sum_k a_{ik} \le F$, $\sum_k a_{ki} \le F$, whereby the utilization of the fabric is maximized.

9. The method according to Claim 8 wherein at each stage only cells that have arrived in the same frame are transmitted to the next stage, wherein $F=D/3T_c$ or $F=D/4T_c$ if cells are reordered at the outputs, wherein *D* is the maximum tolerable delay and T_c is cell time slot duration.

10. The method according to Claim 2 wherein sets are divided in a way that

$$N_f \le (S-U)\cdot D/T_c,$$

where *S* is switching fabric speedup, *U* is targeted utilization of the switching fabric, *D* is the maximum tolerable delay and T_c is cell time slot duration.

11. The method according to Claim 2 further comprising:

grouping cell time slots into frames of length *F*,

allowing in each frame input port (i) to transmit a_{ij} cells or high-priority cells to output port (*j*), and wherein the number of flows sourced by an input SE or bound for an output SE that are balanced starting from different internal SE differ by at most one, wherein:

$$\sum_k a_{ik} \le \begin{cases} SF - \dfrac{N_f}{2} & F \ge \dfrac{N_f}{S} \\ \dfrac{(SF)^2}{2N_f} & F < \dfrac{N_f}{S} \end{cases}, \quad \sum_k a_{ki} \le \begin{cases} SF - \dfrac{N_f}{2} & F \ge \dfrac{N_f}{S} \\ \dfrac{(SF)^2}{2N_f} & F < \dfrac{N_f}{S} \end{cases}$$

where S is the switching fabric speedup.

12. The method according to Claim 11 wherein at each stage only cells that have arrived in the same frame are transmitted to the next stage, wherein F=D/3T_c or $F=D/4T_c$ if cells are reordered at the outputs, wherein D is the maximum tolerable delay and T_c is cell time slot duration.

13. The method according to Claim 11 wherein:

$$S = \begin{cases} 1 + \dfrac{N_f}{2F} & F \ge \dfrac{N_f}{2} \\ \sqrt{\dfrac{2N_f}{F}} & F < \dfrac{N_f}{2} \end{cases},$$

and wherein :

$$\sum_k a_{ik} \le F, \quad \sum_k a_{ki} \le F,$$

whereby utilization of the fabric is maximized.

14. The method according to Claim 13 wherein at each stage only cells that have arrived in the same frame are transmitted to the next stage, wherein $F=D/3T_c$ or $F=D/4T_c$ if cells are reordered at the outputs, wherein D is the maximum tolerable delay and T_c is cell time slot duration.

15. The method according to Claim 11 wherein the counter of each flow is set at the beginning of each frame, and wherein the counter is set to $c_{ij}=(i+j)$ mod l, wherein i is the numerical identifier of an associated input subset and j is the numerical identifier of an associated output subset, comprising the following steps:

(a) transmitting a cell from the flow through the internal switching element associated with a numerical identifier which is equal to the numerical value of the counter of this flow; and

(b) changing the numerical value of the counter by decrementing or incrementing the counter modulus *l*, and

(c) stopping if there are no more cells from the flow, otherwise go to step (a),

wherein the sub-steps are performed for each flow.

16. The method according to Claim 2 wherein the numbers of flows sourced by an input SE or bound for an output SE that are balanced starting from different internal SEs differ by at most 1, wherein the subsets are grouped so that N_f fulfills:

$$N_f \leq \begin{cases} 2(S-U)\cdot F & U \geq \frac{S}{2} \\ \frac{S^2 F}{2U} & U < \frac{S}{2} \end{cases}$$

where *S* is the switching fabric speedup, *U* is targeted utilization of the switching fabric, *D* is the maximum tolerable delay and T_c is cell time slot duration.

17. An article of manufacture for balancing unicast or multicast flows in a multistage non-blocking fabric, wherein the fabric comprises at least one internal switching element (SE) stage, wherein the stage has *l* internal switching elements, wherein each internal switching element is associated with a unique numerical identifier, and wherein the fabric comprises an input SE stage and an output SE stage, wherein the article comprises:

a machine readable medium containing one or more programs which when executed implement the steps of:

(a) dividing input ports into input sets whereby each input set consists of input ports that transmit through the same input SE, and wherein the input sets are further divided into input subsets, and

(b) dividing output ports into output sets whereby each output set consists of output ports that receive cells through the same output SE, and wherein the output sets are further divided into output subsets,

(c) assigning each cell received into the fabric to a flow comprising:

(i) if a cell is a unicast cell, then associating the cell received into the fabric with an input subset and with an output subset based on the input port i and the output port j of the cell, wherein cells which are associated with the same input subset and associated with the same output subset are assigned to the same flow, else

(ii) if a cell is a multicast cell, then the cell is associated with an input subset and associated with multiple output subsets based on the input port i and the multiple output ports $\{j\}$ of the cell, wherein cells which are associated with the same input subset and associated with the output subsets of the same output sets are assigned to the same flow, and

(d) transmitting flows through the internal SE stage wherein cells of a particular flow are distributed among the internal switching elements, wherein the quantity of the cells of each particular flow transmitted at each internal SE differs by at most h, wherein h is positive,

wherein the number of subsets of at least one input set or at least one output set is less than n, wherein n is the number of ports of that input SE or of that output SE, and wherein N is the total number of input ports and output ports, and wherein N_f, is the maximum number of flows whose cells pass any given link, and wherein n, N, N_f, h, i, j and l are natural numbers.

18. The article according to Claim 16 wherein h is one.

19. The article according to Claim 18 wherein (d) comprises a machine readable medium containing one or more programs which when executed implement the steps of:

(a) associating a unique counter with a flow, wherein a counter is designated as c_{ij}, wherein i is the numerical identifier of an associated input subset and j is the numerical identifier of an associated output subset;

(b) initializing the counter with a number less than or equal to l;

(c) transmitting a cell from the flow through the internal switching element associated with a numerical identifier which is equal to the numerical value of the counter; and

(d) changing the numerical value of the counter by decrementing or incrementing the counter modulus *l*, and

(e) stopping if there are no more cells from the flow, otherwise go to step (c),

wherein the sub-steps are performed for each flow.

20. An apparatus for balancing unicast or multicast flows in a multistage non-blocking fabric, wherein the fabric comprises at least one internal switching element (SE) stage, wherein the stage has *l* internal switching elements, wherein each internal switching element is associated with a unique numerical identifier, and wherein the fabric comprises an input SE stage and an output SE stage, the apparatus comprises:

a flow control device configured to:

(a) divide input ports into input sets whereby each input set consists of input ports that transmit through the same input SE, and wherein the input sets are further divided into input subsets, and

(b) divide output ports into output sets whereby each output set consists of output ports that receive cells through the same output SE, and wherein the output sets are further divided into output subsets,

(c) assign each cell received into the fabric to a flow comprising:

(i) if a cell is a unicast cell, then associate the cell received into the fabric with an input subset and with an output subset based on the input port *i* and the multiple output ports {*j*} of the cell, wherein cells which are associated with the same input subset and associated with the same output subsets are assigned to the same flow, else

(ii) if a cell is a multicast cell, then the cell is associate with an input subset and associated with multiple output subsets based on the input port *i* and the multiple output ports {*j*} of the cell, wherein cells which are associated with the same input subset and associated with the output subsets of the same output sets are assigned to the same flow, and

(d) transmit flows through the internal SE stage wherein cells of a particular flow are distributed among the internal switching elements, wherein the quantity of the cells of each particular flow transmitted at each internal SE differs by at most h, wherein h is positive,

wherein the number of subsets of at least one input set or at least one output set is less than n, wherein n is the number of ports of that input SE or of that output SE, and wherein N is the total number of input ports and output ports, and wherein N_f is the maximum number of flows whose cells pass any given link, and wherein n, N, N_f, h, i, j and l are natural numbers.

21. The apparatus according to Claim 20 wherein (d) comprises

a counter module configured to:

(a) associate a unique counter with a flow, wherein a counter is designated as c_{ij};

(b) initialize the counter with a number less than or equal to l;

(c) transmit a cell from the flow through the internal switching element associated with a numerical identifier which is equal to the numerical value of the counter;

(d) change the numerical value of the counter by decrementing or incrementing the counter modulus l, and

(e) stop if there are no more cells from the flow, otherwise go to step (c), wherein the sub-steps are performed for each flow.

22. A multistage non-blocking switch comprising:

(a) at least one internal switching element (SE) stage, wherein the stage has l internal switching elements, wherein each internal switching element is associated with a unique numerical identifier,

(b) an input SE stage,

(c) an output SE stage,

(d) input ports which are divided into input sets wherein each input set consists of input ports that transmit through the same input SE, and wherein the input sets are further divided into input subsets, and

(e) output ports which are divided into output sets wherein each output set consists of output ports that receive cells through the same output SE, and wherein the output sets are further divided into output subsets, and

(f) a flow assignment module wherein the module assigns cells which are received into the fabric to a flow, wherein the assignment comprises

(i) if a cell is a unicast cell, then the cell is associated with an input subset and associated with an output subset based on the input port *i* and the output port *j* of the cell, wherein cells which are associated with the same input subset and associated with the same output subset are assigned to the same flow, else

(ii) if a flow is a multicast flow, then each cell is associated with an input subset and associated with multiple output subsets based on the input port *i* and the multiple output ports $\{j\}$ of the cell, wherein cells which are associated with the same input subset and associated with the output subsets of the same output sets are assigned to the same flow,

whereby flows are transmitted through the internal SE stage wherein cells of a particular flow are distributed among the internal switching elements, wherein the quantity of the cells of each particular flow transmitted at each internal SE differs by at most *h*, wherein *h* is positive,

wherein the number of subsets of at least one input set or at least one output set is less than *n*, wherein *n* is the number of ports of that input SE or of that output SE, and wherein *N* is the total number of input ports and output ports, and wherein N_f, is the maximum number of flows whose cells pass any given link, and wherein *n*, *N*, N_f, *h, i, j* and *l* are natural numbers.

23. The fabric according to Claim 22 wherein the assignment module comprises a lookup table.

24. The fabric according to Claim 22 wherein the assignment module

(a) associates a unique counter with a flow, wherein a counter is designated as c_{ij}, wherein i is the numerical identifier of an associated input subset and j is the numerical identifier of an associated output subset;

(b) initializes the counter with a number less than or equal to l;

(c) transmits a cell from the flow through the internal switching element associated with a numerical identifier which is equal to the numerical value of the counter;

(d) changes the numerical value of the counter by decrementing or incrementing the counter modulus l, and

(e) stops if there are no more cells from the flow, otherwise go to step (c), wherein the sub-steps are performed for each flow.

25. A method for balancing unicast or multicast flows in a multistage non-blocking fabric, wherein the fabric comprises at least one internal switching element (SE) stage, an input SE stage and an output SE stage, wherein the method comprises:

(a) receiving cells into the fabric wherein each cell is associated with an input subset and associated with an output subset according to the source and destination address of the cell,

(b) assigning each cell to a flow, wherein cells sourced from the same input subset, and bound for the same output subset, or multiple output subsets, are assigned to the same flow, and

(c) transmitting flows through the internal SE stage wherein cells of a particular flow are distributed among the internal switching elements, wherein the cells of each particular flow transmitted at each internal SE differs by at most h, wherein h is positive,

whereby the flow in the fabric is balanced.